

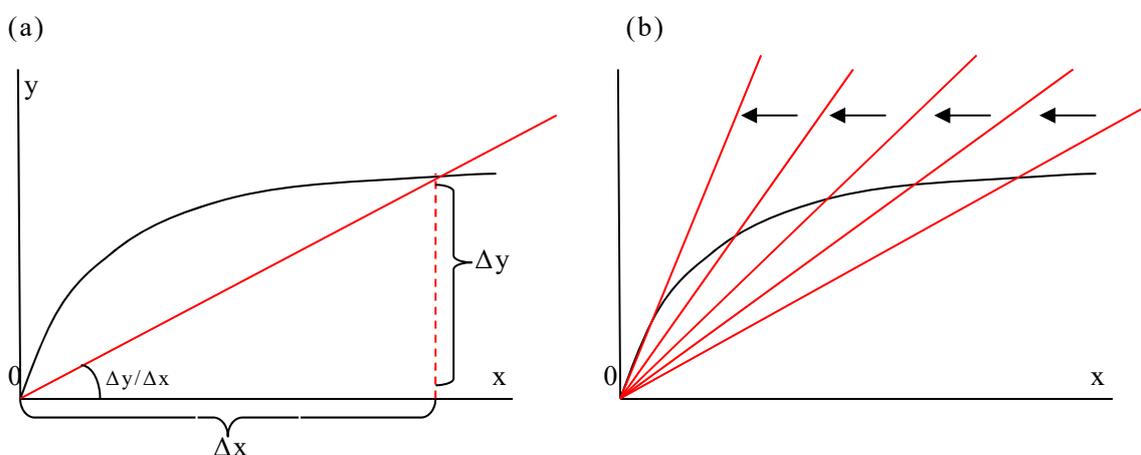
微分について

ある値 x に対して、ある値 $f(x)$ を対応させる仕方を 1 変数の関数といいます。たとえば、 $f(x)=4+2x$ 、 $f(x)=3x^2$ は、変数 x を $f(x)$ に対応させる仕方を示しており、1 変数の関数の例となります。

微分とは、関数 $f(x)$ の接線の傾きであり、 $f'(x)$ や $df(x)/dx$ と表記されます。換言すると、微分は「 x を微小に変化させたとき、 $f(x)$ がどれだけ変化するか」を表します。

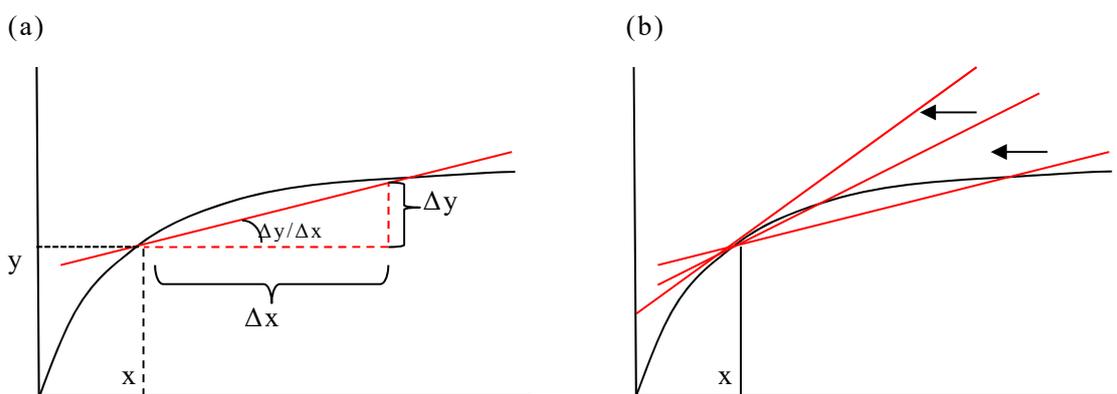
ここで $y=f(x)$ としましょう。図 A-1(a) では、 $y=f(x)$ を描いています。図をみると、 x が増えると y も増加しますが、その増加幅は少しずつ減少する関数であることが分かります。ここで原点 0 から、 x は Δx だけ変化すると、 y は Δy だけ変化するのが分かります（ Δx は x の変化量、 Δy は y の変化量を表す記号と考えてください）。そして、図 A-1(a) において、 Δx を底辺、 Δy を高さとした三角形の傾きは、 $\Delta y/\Delta x$ となります。ここで図 A-1(b) のように、底辺 Δx をどんどん小さくしていくと、三角形の傾き $\Delta y/\Delta x$ は、原点 0 で評価した接線（関数 $f(x)$ の傾き）に近づいていくことが分かります。そして、 Δx が 0 に近いとき、 $\Delta y/\Delta x$ は原点 0 で評価した関数の接線の傾きになります。これが微分となります。

図 A-1 原点 0 での微分のイメージ



もちろん、関数の接線の傾きは、 x をどの値で評価するかで異なるでしょう。たとえば、 x は 0 ではないとしましょう。図 A-2(a)から、ある値 x ($\neq 0$)から $x + \Delta x$ に変化したとき、 y は $y + \Delta y$ に変化する事が分かります。ここで $\Delta y / \Delta x$ は、 Δx が大きいため、関数の傾きにはなっていません。しかし、図 A-2(b)のように、 Δx を 0 に近づけていくと、 $\Delta y / \Delta x$ は、 x における接線の傾きに近づいていきます。これが x で評価した微分となります。

図 A-2 一般的な微分のイメージ



微分のイメージが理解できたでしょうか。繰り返しになりますが、微分とは、関数 $f(x)$ の接線の傾きであり、「 x が変化したとき $y=f(x)$ がどれぐらい変化するか」を表します。数式を用いると、微分は

$$\begin{aligned} \frac{dy}{dx} &= \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \end{aligned}$$

と定義できます。ここで、 $\lim_{\Delta x \rightarrow 0}$ とは x の変化が非常に小さい (Δx が 0 に非常に近い) 状態を表します。

微分の簡単な公式

ここで $y=x^2$ という関数を考えましょう。このとき、微分は $2x$ となります。

$$y=x^2 \quad \Rightarrow \quad \frac{dy}{dx} = 2x$$

[証明] いま x から $x + \Delta x$ に微小に変化したとします。このとき、 $\Delta y / \Delta x$ は、

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = \frac{(x + \Delta x)^2 - x^2}{\Delta x}$$

$$= \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x} = 2x + \Delta x$$

となります。ここで Δx を 0 に近づけると、 $2x$ となるのが分かります。 [終]

一般的には、 $y=x^n$ という関係を考えると、微分は

$$y=x^n \Rightarrow \frac{dy}{dx} = nx^{n-1}$$

となります。

[証明] ここで x から $x+\Delta x$ に変化した状況を考えます。表記を簡単にするため、 $x_*=x+\Delta x$ と定義しましょう (つまり、 $\Delta x=x_*-x$)。よって、微分の定義から、

$$\frac{f(x+\Delta x) - f(x)}{\Delta x} = \frac{x_*^n - x^n}{x_* - x}$$

となります。

ここで右辺の別表現を考えましょう。まず、 $n=2$ なら

$$(x_*^2 - x^2)/(x_* - x) = x_* + x$$

となり、 $n=3$ なら

$$(x_*^3 - x^3)/(x_* - x) = x_*^2 + x_*x + x^2$$

となり、 $n=4$ なら

$$(x_*^4 - x^4)/(x_* - x) = x_*^3 + xx_*^2 + x^2x_* + x^3$$

となります (両辺を $(x_* - x)$ で掛けて、両辺が等式となることを確認してください)。これを一般化すると、任意の n に対して、

$$(x_*^n - x^n)/(x_* - x) = x_*^{n-1} + xx_*^{n-2} + x^2x_*^{n-3} + \dots + x^{n-2}x_* + x^{n-1}$$

と表現できます。ただし、上式において x_*^{n-j} の乗数 $n-j$ が負なら、その項は全て無視します。

この関係を用いると、

$$\frac{f(x+\Delta x) - f(x)}{\Delta x} = \frac{x_*^n - x^n}{x_* - x} = x_*^{n-1} + xx_*^{n-2} + x^2x_*^{n-3} + \dots + x^{n-2}x_* + x^{n-1}$$

となります。ここで Δx を 0 に近づけると、 $x_*=x+\Delta x$ は、 x に限りなく近づきますから、微分は、上式の x_* を x で置き換えた式となります

$$\begin{aligned} & x^{n-1} + xx^{n-2} + x^2x^{n-3} + \dots + x^{n-1} \\ & = x^{n-1} + x^{n-1} + x^{n-1} + \dots + x^{n-1} = nx^{n-1} \end{aligned}$$

[終]

ここで、さらに上の公式を一般化しましょう。任意の定数 a と b を考えて、関数 $y=a+bx^n$ を定義します。この関数を微分すると

$$y=a+bx^n \Rightarrow \frac{dy}{dx} = bnx^{n-1}$$

となります。ここで定数 a は、 x に依存していませんから、微分の式には表れていません。

[証明] x から $x_*=x+\Delta x$ に変化した状況を考えます（つまり、 $\Delta x=x_*-x$ ）。このとき、

$$\frac{f(x+\Delta x)-f(x)}{\Delta x} = \frac{(a+bx_*^n)-(a+bx^n)}{x_*-x} = b \left(\frac{x_*^n-x^n}{x_*-x} \right)$$

ここで $\Delta x=x_*-x$ を 0 に近づけると、 $(x_*^n-x^n)/(x_*-x)$ は nx^{n-1} になりますから、微分は bnx^{n-1} となります。[終]

例 1 : $y=a$ としましょう。このとき、 a は x に依存していませんから、

$$dy/dx = 0$$

となります。先の公式で $b=0$ とした場合に該当します。定数の微分は 0 となります。

例 2 : $y=bx$ としましょう。このとき、公式より、

$$dy/dx = bx^0 = b$$

となります。つまり、 x で微分すると定数 b となります。

例 3 : $y=3+4x^3$ としましょう。このとき、公式より、

$$dy/dx = 4 \times 3x^2 = 12x^2$$

となります。

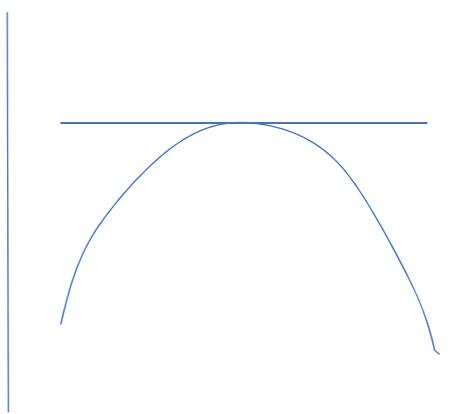
関数の最大化、最小化

微分を用いることで何が分かるのでしょうか。微分は関数の傾き、つまり x が変化したときの y の変化を教えてくれるので、それ自体で有用な情報を与えてくれます。それ以外にも、微分を用いることで、関数 $f(x)$ の最大値や最小値を簡単に見つけることができます。

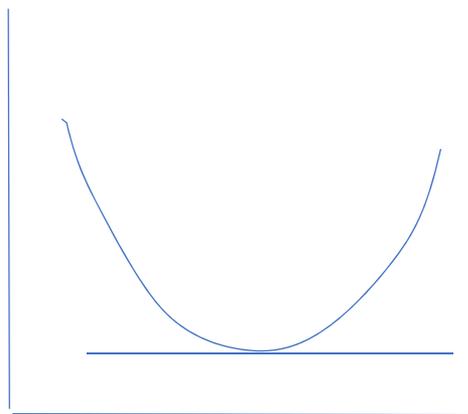
図 A-3 では、2 つの関数 $f(x)$ を描いています。これをみると、関数の最大値と最小値において、関数の傾きが 0 になっているのがみてとれるでしょう。実は、関数 $f(x)$ の微分が 0 となるポイント x とは、関数を最大化もしくは最小化する x になっているのです。

図 A-3 微分が 0 となるポイント

(a) 最大値



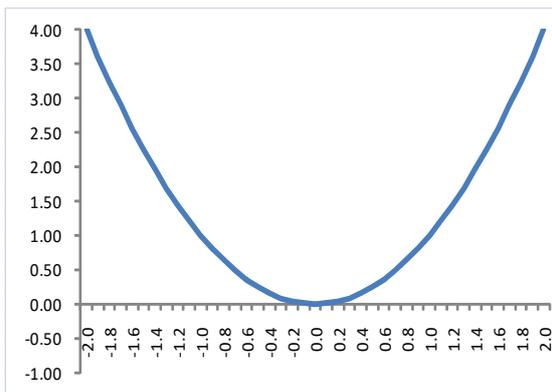
(b) 最小値



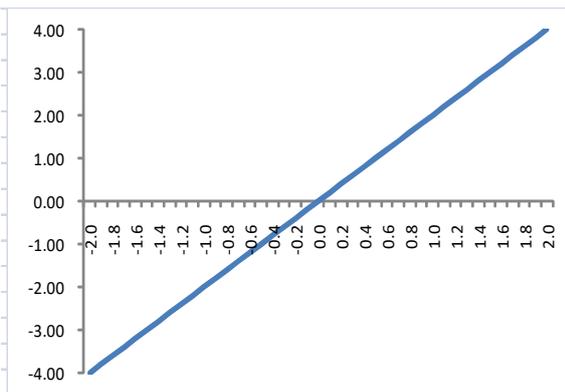
例 1 ここで $y=x^2$ としましょう。このとき、公式より $dy/dx=2x$ となります。図 A-4(a)には、 $y=x^2$ という関数を描いています。また、図 A-4(b)には、微分（接線の傾き）である $2x$ を描きました。当然ですが、 $y=x^2$ という関数は、 x が負のとき、 x が増えると y の値が小さくなり、 x が正のとき、 x が増えると y の値が大きくなります。これは $x < 0$ で関数の傾きが負になり、 $x > 0$ で傾きが正となることを意味します。傾きが正ということは、 x が増えれば y も増えるということです。

図 A-4 関数 $y=x^2$

(a)



(b)



$x=0$ で、傾きは 0 となり、この関数は最小値となります（微分は $2x$ ですから、 $2x=0$ を満たす x は、 $x=0$ となります）。

これまでの議論から、関数を微分して 0 と置いた式を x について解くことで、関数を最大化もしくは最小化するポイント x^* を求められることが理解できました。では、こうして求めた点 x^* が、最大化点か最小化点か、どうしたら判断できるのでしょうか。もちろん関数を図示すれば分かりますが、図に描くのも面倒です。別の方法として、微分して 0 と置いた点 x^* から、微小に x を変化させて、関数の値が減少（増加）すれば、 x^* は最大（最小）化点と判断できます。また、2 階微分¹をみることも、微分が 0 となる x^* が、最大化するポイントか、最小化するポイントかを判断できます。2 階微分の話は少し難しいので、興味がある人は数学の本を読んでみてください。

合成関数の微分

ここで $y=f(x)$ とし、また x は別の変数 z の関数としましょう（これを $g(z)$ と表しましょう）。このとき、 y は z の関数となっており、 $f(g(z))$ を $f(x)$ と $g(z)$ の合成関数と呼びます。たとえば、 $y=(3+2z)^2$ とすると、これは $y=x^2$ （つまり、 $f(x)=x^2$ ）、 $x=3+2z$ （つまり、 $g(z)=3+2z$ ）という合成関数と考えることができます。

合成関数の微分は、関数 $f(g(z))$ の傾きとなります。これは連鎖公式(チェーン・ルール)によって導くことができます。

$$\frac{dy}{dz} = \frac{dy}{dx} \frac{dx}{dz}$$

つまり、 $y=f(g(z))$ の微分は、 $y=f(x)$ の微分を $x=g(z)$ の微分でかけた値となります。換言すれば、連鎖公式は

$$\begin{aligned} (\text{z が変化したときの y の変化量}) &= (\text{z が変化したときの x の変化量}) \\ &\quad \times (\text{x が変化したときの y の変化量}) \end{aligned}$$

となるとしています。これは直観的な式ではないでしょうか。

¹ 微分をさらに微分すること（関数の傾きを微分すること）、換言すれば、「 x が増加したとき、関数の傾きがどれぐらい変化するか」を表します。

例1 関数として、 $y=(3+2z)^2$ とすると、これは $y=x^2$ 、 $x=3+2z$ という合成関数と考えることができます。したがって、連鎖公式から

$$\frac{dy}{dz} = \frac{dy}{dx} \frac{dx}{dz} = (2x)(2) = 4(3+2z)$$

となります。

例2 ここで $y=(a+bz^n)^m$ とします。これは $y=x^m$ 、 $x=a+bz^n$ と定義すれば、連鎖公式から、

$$\begin{aligned} \frac{dy}{dz} &= \frac{dy}{dx} \frac{dx}{dz} = (mx^{m-1})(bnz^{n-1}) \\ &= (m(a+bz^n)^{m-1})(bnz^{n-1}) \end{aligned}$$

となります。

偏微分

これまで、ある変数 x が、ある値 $f(x)$ に対応する 1 変数の関数を考えていました。しかし、もし 2 変数 x_1 、 x_2 が、ある値 $f(x_1, x_2)$ に対応するならば、これは 2 変数の関数といえます。2 つ以上の変数の関数も同様に定義できます。

ここで x_1 に関する偏微分とは、 x_2 を固定した値と考えて、 x_1 に関して微分をとったものとなります。これは

$$\begin{aligned} \frac{\partial y}{\partial x_1} &= \lim_{\Delta x_1 \rightarrow 0} \frac{\Delta y}{\Delta x_1} \\ &= \lim_{\Delta x_1 \rightarrow 0} \frac{f(x_1 + \Delta x_1, x_2) - f(x_1, x_2)}{\Delta x_1} \end{aligned}$$

となります。ここで、 x_2 を固定して、 x_1 だけが $x_1 + \Delta x_1$ だけ変化していることに注意してください。

偏微分では、微分と明確に区別するため、記号として d ではなく ∂ (ラウンド・デルタと呼ぶ) を用いています。偏微分記号を用いるときは、変数が複数あり、そのうちの 1 つを微小に変化させる状況を考えているのです。

偏微分と聞くと難しそうですが、他の変数を固定した値とみなして、微分をとるだけなので、とても簡単なことが分かってもらえると思います。

例 1 ここで $y=x_1^2x_2$ を、 x_1 で偏微分してみましょう。ここで x_2 は固定した値とみなします。よって、偏微分は以下となります。

$$\frac{\partial y}{\partial x_1} = 2x_1x_2$$

例 2(残差 2 乗和) 10 章では、最小 2 乗推定量を求める際、残差 2 乗和を最小化する α と β を求めました。残差は $u=(y-\alpha-\beta x)$ であり、残差 2 乗 $u^2=(y-\alpha-\beta x)^2$ となります。ここで残差 2 乗和を α と β で偏微分するとどうなるかをみてみましょう。その際、 y と x は固定した値であり、 α と β だけが変数と考えましょう(つまり、関数 u^2 は α と β だけに依存しています)。まず、関数 $u^2=(y-\alpha-\beta x)^2$ を α で偏微分すると、

$$\frac{\partial u^2}{\partial \alpha} = \frac{\partial u^2}{\partial u} \frac{\partial u}{\partial \alpha} = (2u)(-1) = -2(y - \alpha - \beta x)$$

となります。ここで連鎖公式を使いました。また、 β で偏微分すると、

$$\frac{\partial u^2}{\partial \beta} = \frac{\partial u^2}{\partial u} \frac{\partial u}{\partial \beta} = (2u)(-x) = -2(y - \alpha - \beta x)x$$

となります。次に、残差 2 乗和

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

を偏微分してみましょう。これは単に、 α と β で各項を偏微分するわけですから、簡単に計算できます。つまり、

$$\begin{aligned} \frac{\partial \sum_{i=1}^n u_i^2}{\partial \alpha} &= \sum_{i=1}^n \frac{\partial u_i^2}{\partial \alpha} = \sum_{i=1}^n \frac{\partial u_i^2}{\partial u_i} \frac{\partial u_i}{\partial \alpha} = \sum_{i=1}^n (2u_i)(-1) = -2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) \\ \frac{\partial \sum_{i=1}^n u_i^2}{\partial \beta} &= \sum_{i=1}^n \frac{\partial u_i^2}{\partial \beta} = \sum_{i=1}^n \frac{\partial u_i^2}{\partial u_i} \frac{\partial u_i}{\partial \beta} = \sum_{i=1}^n (2u_i)(-x_i) = -2 \sum_{i=1}^n (y_i - \alpha - \beta x_i)x_i \end{aligned}$$

となります。10 章で説明した通り、これらの式を 0 と置いて、両式を満たす α と β が最小 2 乗推定量となります。